

FOR OFFICIAL USE ONLY

CRF

ACCESS DB # 174850  
PLEASE PRINT CLEARLY

Scientific and Technical Information Center

SEARCH REQUEST FORM

12/22 02p

Requester's Full Name: DAVID GUZO Examiner #: 70677 Date: 12/21/05  
Art Unit: 1636 Phone Number: 2-0767 Serial Number: 10/696282  
Location (Bldg/Room#): 2A79 (Mailbox #): 2C70 Results Format Preferred (circle) PAPER DISK  
\*\*\*\*\*

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Date: \_\_\_\_\_

Search Topic:  
Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc., if known.

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please run a regular + interference sequence  
search on SEQ ID NO: 15 and 17.

AA 599

AA 534

Thinks

RECEIVED  
DEC 21 2005  
SEARCH/CHIEF, DIVISION  
(STIC)

MEY





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2005, 19:04:59 ; Search time 32.0494 Seconds  
(without alignments)  
1377.523 Million cell updates/sec

Title: US-10-696-282-17

Perfect score: 2906  
Sequence: 1 MASGGAPADNNEGADGVG.....NNGLYTERPPICTRYLTRPL 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Parents AA:  
1: /cgn2\_6/prodata/1/iaa/5.COMB.pep.\*  
2: /cgn2\_6/prodata/1/iaa/6.COMB.pep.\*  
3: /cgn2\_6/prodata/1/iaa/H.COMB.pep.\*  
4: /cgn2\_6/prodata/1/iaa/PCITUS.COMB.pep.\*  
5: /cgn2\_6/prodata/1/iaa/RE.COMB.pep.\*  
6: /cgn2\_6/prodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2906	100.0	534	2	US-09-807-802A-17	Sequence 17, Appl
2	2906	100.0	599	2	US-09-807-802A-15	Sequence 15, Appl
3	2906	100.0	736	2	US-09-807-802A-3	Sequence 3, Appl
4	2906	100.0	736	2	US-09-807-802A-13	Sequence 13, Appl
5	2481.5	85.4	533	2	US-10-038-972A-15	Sequence 15, Appl
6	2481.5	85.4	598	2	US-10-038-972A-14	Sequence 14, Appl
7	2481.5	85.4	735	2	US-09-321-589-1	Sequence 1, Appl
8	2481.5	85.4	735	2	US-10-293-478-1	Sequence 1, Appl
9	2481.5	85.4	735	2	US-10-038-972A-13	Sequence 13, Appl
10	1690.5	58.2	544	2	US-09-532-594B-18	Sequence 18, Appl
11	1690.5	58.2	598	2	US-09-532-594B-16	Sequence 16, Appl
12	1690.5	58.2	734	2	US-09-532-594B-4	Sequence 4, Appl
13	1665	57.3	532	2	US-09-533-427-6	Sequence 6, Appl
14	1665	57.3	588	2	US-09-533-427-5	Sequence 5, Appl
15	1665	57.3	724	2	US-09-533-427-4	Sequence 4, Appl
16	491.5	16.9	756	2	US-09-438-268-4	Sequence 29, Appl
17	486	16.7	554	2	US-10-187-253E-29	Sequence 29, Appl
18	486	16.7	554	2	US-10-187-253E-35	Sequence 35, Appl
19	486	16.7	781	2	US-10-187-253E-27	Sequence 27, Appl
20	479.5	16.5	543	2	US-08-856-841-22	Sequence 22, Appl
21	477	16.4	781	2	US-10-187-253E-33	Sequence 33, Appl
22	326	11.2	415	2	US-08-856-841-20	Sequence 20, Appl
23	318	10.9	395	2	US-08-856-841-13	Sequence 13, Appl
24	318	10.9	486	2	US-08-856-841-19	Sequence 19, Appl
25	317.5	10.9	500	2	US-08-856-841-16	Sequence 16, Appl
26	317.5	10.9	501	2	US-08-856-841-18	Sequence 18, Appl
27	304.5	10.5	264	2	US-08-856-841-14	Sequence 14, Appl

28	303	10.4	398	2	US-08-856-841-21	Sequence 21, Appl
29	209.5	7.2	579	6	5223424-13	Patent No. 5223424
30	183	6.3	584	2	US-09-022-949-2	Sequence 2, Appl
31	171	5.9	387	2	US-08-856-841-17	Sequence 17, Appl
32	132.5	4.6	489	2	US-10-376-397B-4	Sequence 4, Appl
33	119	4.1	1095	2	US-09-107-532A-3855	Sequence 3855, Ap
34	111.5	3.8	551	2	US-09-246-786A-23413	Sequence 23413, A
35	111	3.8	655	1	US-08-469-202-27	Sequence 27, Appl
36	111	3.8	655	1	US-08-484-434C-34	Sequence 34, Appl
37	111	3.8	655	2	US-09-384-361-34	Sequence 4, Appl
38	109	3.8	1186	1	US-08-485-568A-4	Sequence 6, Appl
39	109	3.8	1186	1	US-08-357-698-6	Sequence 4, Appl
40	109	3.8	1186	1	US-08-590-554A-4	Sequence 4, Appl
41	109	3.8	1186	1	US-09-184-223-4	Sequence 4, Appl
42	109	3.8	1186	4	PCR-US93-12682-6	Sequence 6, Appl
43	108.5	3.7	1637	2	US-10-172-502-14	Sequence 14, Appl
44	108	3.7	624	2	US-08-947-965-78	Sequence 78, Appl
45	108	3.7	655	1	US-08-469-202-28	Sequence 28, Appl

## ALIGNMENTS

RESULT 1  
US-09-807-802A-17  
Sequence 17, Application US/09807802A  
Patent No. 6759237  
GENERAL INFORMATION:  
APPLICANT: Watson, James M.  
TITLE OF INVENTION: Adeno-Associated Virus Serotype I Nucleic Acid Sequences,  
TITLE OF INVENTION: Vectors and Host Cells Containing Same  
FILE REFERENCE: GNVN.031USA  
CURRENT APPLICATION NUMBER: US/09/807,802A  
CURRENT FILING DATE: 2002-02-21  
PRIOR APPLICATION NUMBER: US 60/107,114  
PRIOR FILING DATE: 1998-11-05  
PRIOR APPLICATION NUMBER: PCT/US99/25694  
PRIOR FILING DATE: 1999-11-02  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 17  
LENGTH: 534  
TYPE: PRT  
ORGANISM: AAV-1  
US-09-807-802A-17

Query Match 100.0%; Score 2906; DB 2; Length 534;  
Best Local Similarity 100.0%; Pred. No. 9.6e-258;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MASGGAPADNNEGADGVGNASGNHCDSTWLGDRVITTTSTRTWALPTYYNNHLYQIIS	60
DB	1	MASGGAPADNNEGADGVGNASGNHCDSTWLGDRVITTTSTRTWALPTYYNNHLYQIIS	60
QY	61	ASTGASNDNHYFEGYSPMGYFDFNRFCHSPSPDQRLINNNMGFPKRLNPLFNIQYK	120
DB	61	ASTGASNDNHYFEGYSPMGYFDFNRFCHSPSPDQRLINNNMGFPKRLNPLFNIQYK	120
QY	121	EVTINDGVTTIANLNTSTVOVFSDSSEYQLPYVLGSAHOCCLPPFADVFMIDPYGYLTIN	180
DB	121	EVTINDGVTTIANLNTSTVOVFSDSSEYQLPYVLGSAHOCCLPPFADVFMIDPYGYLTIN	180
QY	181	NGSQAVRRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVPHSHSYAHSCSLDRMLNPLIDY	240
DB	181	NGSQAVRRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVPHSHSYAHSCSLDRMLNPLIDY	240
QY	241	LYTINRQNGSGAOKKDLFSRGSPPAGMSVOPKMLPGCYRQOQSVSTKTDNNNSNFT	300
DB	241	LYTINRQNGSGAOKKDLFSRGSPPAGMSVOPKMLPGCYRQOQSVSTKTDNNNSNFT	300
QY	301	WTGASKYLNLRGSIINPTAMASHKDDDEKFFPMGVMIFGKESAGASTALDNYMTD	360
DB	301	WTGASKYLNLRGSIINPTAMASHKDDDEKFFPMGVMIFGKESAGASTALDNYMTD	360

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2005, 18:55:18 ; Search time 139.509 Seconds  
(without alignments)  
2700.553 Million cell updates/sec

Title: US-10-696-282-17

Perfect score: 2906

Sequence: 1 MASGGAPMADNNEGADGVC.....NNGLYTERPRIGTRVLTPL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2906	100.0	736	Q9WB88_VVIRU	Q9WB88 adeno-ssoc
2	2892	99.5	736	Q6JC12_VVIRU	Q6JC12 adeno-ssoc
3	2885	99.3	736	Q6JC08_VVIRU	Q6JC08 adeno-ssoc
4	2884	99.2	736	Q56137_VVIRU	Q56137 adeno-ssoc
5	2883	99.2	737	Q6JC13_VVIRU	Q6JC13 adeno-ssoc
6	2863	98.5	736	Q6JC10_VVIRU	Q6JC10 adeno-ssoc
7	2536.5	87.3	735	Q6JC34_VVIRU	Q6JC34 adeno-ssoc
8	2535.5	87.3	735	Q6JC08_VVIRU	Q6JC08 adeno-ssoc
9	2532.5	87.1	735	Q6JC28_VVIRU	Q6JC28 adeno-ssoc
10	2530.5	87.1	735	Q6JC27_VVIRU	Q6JC27 adeno-ssoc
11	2529.5	87.0	736	Q56139_VVIRU	Q56139 adeno-ssoc
12	2529.5	87.0	735	Q6JC21_VVIRU	Q6JC21 adeno-ssoc
13	2528	87.0	735	Q6JC11_VVIRU	Q6JC11 adeno-ssoc
14	2524.5	86.9	735	Q6JC38_VVIRU	Q6JC38 adeno-ssoc
15	2524.5	86.9	735	Q6JC42_VVIRU	Q6JC42 adeno-ssoc
16	2520.5	86.7	735	Q6JC17_VVIRU	Q6JC17 adeno-ssoc
17	2517	86.6	734	Q6JC04_VVIRU	Q6JC04 adeno-ssoc
18	2515.5	86.6	735	Q6JC44_VVIRU	Q6JC44 adeno-ssoc
19	2514.5	86.5	735	Q6JC36_VVIRU	Q6JC36 adeno-ssoc
20	2513.5	86.5	735	Q6JC26_VVIRU	Q6JC26 adeno-ssoc
21	2511.5	86.4	734	Q6JC02_VVIRU	Q6JC02 adeno-ssoc
22	2506.5	86.3	735	Q6JC43_VVIRU	Q6JC43 adeno-ssoc
23	2502.5	86.1	735	Q6JC20_VVIRU	Q6JC20 adeno-ssoc
24	2500.5	86.0	735	Q6JC25_VVIRU	Q6JC25 adeno-ssoc
25	2499.5	86.0	735	Q6JC04_VVIRU	Q6JC04 adeno-ssoc
26	2493.5	85.8	735	Q6JC05_VVIRU	Q6JC05 adeno-ssoc
27	2493.5	85.8	735	Q6JC08_VVIRU	Q6JC08 adeno-ssoc
28	2493.5	85.8	735	Q6JC29_VVIRU	Q6JC29 adeno-ssoc
29	2492	85.8	734	Q6JC03_VVIRU	Q6JC03 adeno-ssoc
30	2491.5	85.7	735	Q6JC41_VVIRU	Q6JC41 adeno-ssoc
31	2490.5	85.7	735	Q6JC06_VVIRU	Q6JC06 adeno-ssoc

32	2490.5	85.7	735	Q6JC08_VVIRU	Q6JC08 adeno-ssoc
33	2488.5	85.6	735	Q6JC07_VVIRU	Q6JC07 adeno-ssoc
34	2488.5	85.6	735	Q6JC24_VVIRU	Q6JC24 adeno-ssoc
35	2488.5	85.6	735	Q6JC39_VVIRU	Q6JC39 adeno-ssoc
36	2484.5	85.5	735	Q6JC01_VVIRU	Q6JC01 adeno-ssoc
37	2483.5	85.5	735	Q6JC01_VVIRU	Q6JC01 adeno-ssoc
38	2483.5	85.5	735	Q6JC25_VVIRU	Q6JC25 adeno-ssoc
39	2483.5	85.5	735	Q6JC35_VVIRU	Q6JC35 adeno-ssoc
40	2481.5	85.4	733	Q92917_AAV2	Q92917 adeno-ssoc
41	2481.5	85.4	598	Q56652_AAV2	Q56652 adeno-ssoc
42	2481.5	85.4	735	Q56652_AAV2	Q56652 adeno-ssoc
43	2480.5	85.4	735	Q6JC31_VVIRU	Q6JC31 adeno-ssoc
44	2478.5	85.3	736	Q6JC24_VVIRU	Q6JC24 adeno-ssoc
45	2476.5	85.2	735	Q6JC33_VVIRU	Q6JC33 adeno-ssoc

## ALIGNMENTS

RESULT 1	ID	Q9WB88_VVIRU	PRELIMINARY:	PRT:	736 AA.
AC	Q9WB88	Q9WB88_VVIRU	Q9WB88_VVIRU		
DT	01-NOV-1999	(Tremblrel. 12, Created)			
DT	01-NOV-1999	(Tremblrel. 12, Last sequence update)			
DT	01-MAR-2004	(Tremblrel. 26, Last annotation update)			
DE	Capsid protein.				
OS	Adeno-associated virus 1.				
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.				
OX	NCBI_TaxID=85106;				
RN	(1)				
RP	NUCLEOTIDE SEQUENCE.				
RX	MEDLINE=99214338; PubMed=10196295;				
RA	Xiao W., Chirmule N., Berta S.C., McCullough B., Gao G., Wilson J.M.;				
RT	"Gene therapy vectors based on adeno-associated virus type 1."				
RL	J. Virol. 73:3994-4003(1999).				
DR	EMBL; AF063497; AAD2757.1; -; Genomic DNA.				
DR	SMR; Q9WB88; 217-736.				
DR	GO; GO:0019028; Cytical capsid; IEA.				
DR	GO; GO:0005198; F:structural molecule activity; IEA.				
DR	InterPro; IPR001403; Parvo_coat.				
DR	Pfam; PF00740; Parvo_coat; 1.				
SQ	SEQUENCE: 736 AA; 81375 MW; CFAFB9BD5CD0595 CRC64;				
Query Match	100.0%;	Score 2906;	DB 2;	Length 736;	
Best Local Similarity	100.0%;	Pred. No. 1.2e-203;			
Matches 534;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	MASGGAPMADNNEGADGVCNAGMHCDSYGLGRVITTSRTVALPTYNHLYKQISS	60		
DB	203	MASGGAPMADNNEGADGVCNAGMHCDSYGLGRVITTSRTVALPTYNHLYKQISS	262		
QY	61	ASTGASNDNHYFGYSTPGYFDPNRFCHFSRDMQRLINNNGFRPKRLFKFNIOVK	120		
DB	263	ASTGASNDNHYFGYSTPGYFDPNRFCHFSRDMQRLINNNGFRPKRLFKFNIOVK	322		
QY	121	EYTTNDGVTIANNITSTVOVFSDEYQLPYLGSAGHGCCLPPFADVFPMIPQYGLTLN	180		
DB	323	EYTTNDGVTIANNITSTVOVFSDEYQLPYLGSAGHGCCLPPFADVFPMIPQYGLTLN	382		
QY	181	NSGSAVGRSSFYCLFYPSPQMLRTGNFTFSYFEEVFFHSSYASOSLDRLMPLIOY	240		
DB	383	NSGSAVGRSSFYCLFYPSPQMLRTGNFTFSYFEEVFFHSSYASOSLDRLMPLIOY	442		
QY	241	LYTLNRKTONQSSAONKDLFSGSGPAGMSVQPKNMLPGPCYRQGRVSKTKTDNNNSFT	300		
DB	443	LYTLNRKTONQSSAONKDLFSGSGPAGMSVQPKNMLPGPCYRQGRVSKTKTDNNNSFT	502		
QY	301	WTGASRYNLNGRESLIINGTAMASHKDDKDFPMSGVMITGKESAGSNTALDVMITD	360		
DB	503	WTGASRYNLNGRESLIINGTAMASHKDDKDFPMSGVMITGKESAGSNTALDVMITD	562		
QY	361	EEIKAITNPVATERGTVAIVNFPSSSTDPAICGVHAMCALPGMVQDRDVLQGPIMAKI	420		

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2005, 19:04:49 ; Search time 25.451 Seconds  
(without alignments)  
2018.770 Million cell updates/sec

Title: US-10-696-282-17

Perfect score: 2906

Sequence: 1 MASGGAPMADNNEGADGVG.....NNGLYTEPRPIGRTYRLRPL 534

Scoring table: BLOSUM62  
Gap: 10.0, Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2177.5	74.9	504	1 VCPV3A	coat protein - ade
2	1676	57.7	732	2 S52210	coat protein VPI -
3	517	17.8	673	1 VCPVBS	coat protein VPI -
4	487.5	16.8	781	1 VCPV15	coat protein VPI -
5	234.5	8.1	723	1 VCPVPP	coat protein VPI -
6	232.5	8.0	729	1 A60006	coat protein VPI -
7	232.5	8.0	729	1 VCPVNA	coat protein VPI -
8	229	7.9	587	1 B44276	coat protein VPI -
9	226	7.8	722	1 VCPV2	coat protein VPI -
10	197	6.8	716	1 VCPV2M	coat protein VPI -
11	196.5	6.8	716	1 VCPVIM	coat protein VPI -
12	194	6.7	722	1 VCPVME	coat protein VPI -
13	194	6.7	727	1 VCPV1F	coat protein VPI -
14	189	6.5	584	2 S49594	capsid protein VP2
15	189	6.5	727	1 VCPVFP	coat protein VPI -
16	188	6.5	737	1 VCPVCD	coat protein VPI -
17	184	6.3	748	1 VCPVCP	coat protein VPI -
18	177.5	6.1	702	1 VCPVAP	coat protein VPI -
19	175	6.0	722	1 VCPVPCN	coat protein VPI -
20	123.5	4.2	690	2 ABO124	probable TonB-dep
21	122.5	4.2	880	1 SYBSVS	valine-tRNA ligase
22	121	4.2	635	2 P96660	protein F2K1.10 f
23	118	4.1	739	2 T52289	probable transkeo
24	115.5	4.0	345	1 G97024	probable phosphos
25	113.5	3.9	2817	2 B97033	uncharacterized pr
26	109	3.8	418	2 T35753	probable periplasm
27	109	3.8	1072	2 A86827	hypothetical prote
28	109	3.8	1186	2 T18210	delta endotoxin
29	108.5	3.7	1711	2 AB1283	peptidoglycan link

30	108	3.7	555	1 ALK8G	cyclomaldextrin
31	108	3.7	772	1 A44052	outer layer protei
32	108	3.7	826	2 AD1683	penicillin-binding
33	107.5	3.7	413	2 T23098	hypothetical prote
34	107.5	3.7	648	2 S50856	whn protein - rat
35	107	3.7	777	2 AEO462	probable exported
36	107	3.7	1338	2 T30565	MAP kinase kinase
37	106.5	3.7	403	2 B87513	acyl-CoA dehydroge
38	106	3.6	1271	2 D64237	hypothetical prote
39	105.5	3.6	2500	2 G71609	hypothetical prote
40	105	3.6	1394	2 A29637	position-specific
41	104.5	3.6	559	2 S33724	transcription fact
42	104.5	3.6	717	2 H85171	DEF (C1A1) protein
43	104	3.6	395	2 S50986	MAF1 protein - yea
44	104	3.6	715	2 H90977	hypothetical prote
45	104	3.6	719	2 F85824	hypothetical prote

## ALIGNMENTS

### RESULT 1

coat protein - adeno-associated virus type 2

C/Species: adeno-associated virus type 2  
C/Date: 05-Apr-1983 #sequence\_revision 05-Apr-1983 #exc\_change 09-Jul-2004

C/Accession: A03698  
R/Silvastrava, A.; Lueby, E.W.; Berns, K.I.

J. Virol. 45, 555-564, 1983  
A/Title: Nucleotide sequence and organization of the adeno-associated virus 2 genome.

A/Reference number: A03694; MUID: 83164299; PMID: 6300419

A/Accession: A03698  
A/Status: translation not shown

A/Molecule type: DNA

A/Cross-references: UNIPROT: P03135; UNIPARC: UP10000127C8E; EMBL: J01901; NID: G209616; PII

A/Keyfamily: adeno-associated virus coat protein

C/Keywords: coat protein

Query Match 74.9%; Score 2177.5; DB 1; Length 504;  
Best Local Similarity 82.5%; Pred. No. 1.4e-152;  
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;

QY	1	MASGGAPMADNNEGADGVGASGNHCDSTWLCDRVITTSRTWALPTNNHLYKQISS	60
DB	1	MATSGAPMADNNEGADGVGNSGNHCDSTWMDRVTITSTWALPTNNHLYKQISS	60
QY	61	ASTGASNNHYFGYSTPWFGRFDFRPHCHFSPPDMQRLINNMGFRPKLNFQVOK	120
DB	61	OS-GASNNHYFGYSTPWFGRFDFRPHCHFSPPDMQRLINNMGFRPKLNFQVOK	119
QY	121	EYTTNDGVTTANNLTSTVQVFSDESEYOLPVVLASAHQGLPPPADVFM1POYGYLTLN	180
DB	120	EYTTNDGVTTANNLTSTVQVFSDESEYOLPVVLASAHQGLPPPADVFM1POYGYLTLN	179
QY	181	NGSQAVRSFYCLEYFPSPQMLRTGNNFTSFYEEVPHSSVYASOSLDRLNNPLIDQY	240
DB	180	NGSQAVRSFYCLEYFPSPQMLRTGNNFTSFYEEVPHSSVYASOSLDRLNNPLIDQY	239
QY	241	LYTLAKTQNGSQAOKDLFSSRGSBAGMSVQPKWLPGRPYRQORVSKTIDNNNSNFT	300
DB	240	LYTLAKTQNGSQAOKDLFSSRGSBAGMSVQPKWLPGRPYRQORVSKTIDNNNSNFT	299
QY	301	WTGASKYNIANGRESINFGTMAAHKDEDKFFPMGSGVMIFGKESAGASNTALDNW1TD	360
DB	300	WTGASKYNIANGRESINFGTMAAHKDEDKFFPMGSGVMIFGKESAGASNTALDNW1TD	357
QY	361	EEELKAINPVATERRFGVAANVFQSSSTDPAADVHAMGALPGMTWDRDVLQCP1AKI	420
DB	360	EEELKAINPVATERRFGVAANVFQSSSTDPAADVHAMGALPGMTWDRDVLQCP1AKI	417
QY	421	PHTDGHRPSLMGFGKXNPPQILIKNTTVNPNPAEPATYFASFIQYSG	475
DB	420	PHTDGHRPSLMGFGKXNPPQILIKNTTVNPNPAEPATYFASFIQYSG	475

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using bw model

Run on: December 22, 2005, 18:54:53 / Search time 124.898 Seconds  
(without alignments)  
1878.552 Million cell updates/sec

Title: US-10-696-282-17

Sequence: 1 MASGGAPMADNNEGADGVC.....NNGLYTEPRPTGRTYLRPL 534

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 2443163 seqs, 43937871 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

1: Geneseqp21:\*  
2: geneseqp1980s:\*  
3: geneseqp1990s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2906	100.0	534	3	AAV71169
2	2906	100.0	599	3	AAV71168
3	2906	100.0	736	3	AAV71167
4	2906	100.0	736	6	ABB80232
5	2906	100.0	736	7	ABR62762
6	2906	100.0	736	7	ABR62762
7	2906	100.0	736	7	ADV70294
8	2906	100.0	736	9	ADZ27069
9	2906	100.0	736	9	ADZ27012
10	2902	99.9	735	9	ADV67506
11	2892	99.5	736	9	ADZ27008
12	2892	99.5	737	9	ADZ27086
13	2888	99.4	736	9	ADZ27074
14	2885	99.3	736	9	ADZ27007
15	2884	99.2	736	4	ABR59847
16	2884	99.2	736	7	ADV67566
17	2884	99.2	736	9	ADV70293
18	2884	99.2	736	9	ADZ27070
19	2883	99.2	737	9	ADZ27010
20	2863	98.5	736	9	ADZ27009
21	2536.5	87.3	735	9	ADZ26993
22	2532.5	87.1	735	9	ADZ26996
23	2530.5	87.1	735	9	ADZ27034
24	2530	87.1	736	4	ABR59846

25	2529.5	87.0	735	9	ADZ27001	Adz27001 Adeno-sss
26	2529.5	87.0	735	9	ADZ27076	Adz27076 Adeno-sss
27	2529.5	87.0	735	9	ADZ27000	Adz27000 Adeno-sss
28	2528	87.0	736	4	ABR59845	ABR59845 AAV3A cap
29	2528	87.0	736	6	ABR80233	ABR80233 AAV3 VP1
30	2528	87.0	736	7	ABR62763	ABR62763 Adeno-sss
31	2528	87.0	736	7	ADBR6572	ADBR6572 Adeno-sss
32	2528	87.0	736	9	ADV67507	ADV67507 Amino aci
33	2528	87.0	736	9	ADZ27067	Adz27067 Adeno-sss
34	2527.5	87.0	735	9	ADZ26992	Adz26992 Adeno-sss
35	2524.5	86.9	735	9	ADZ27003	Adz27003 Adeno-sss
36	2524.5	86.9	735	9	ADZ26998	Adz26998 Adeno-sss
37	2522.5	86.8	735	9	ADZ27002	Adz27002 Adeno-sss
38	2520.5	86.7	735	9	ADZ26991	Adz26991 Adeno-sss
39	2520.5	86.7	735	9	ADZ27004	Adz27004 Adeno-sss
40	2517	86.6	734	9	ADZ27036	Adz27036 Adeno-sss
41	2515.5	86.6	735	9	ADZ26994	Adz26994 Adeno-sss
42	2514.5	86.5	735	9	ADZ26999	Adz26999 Adeno-sss
43	2513.5	86.5	735	9	ADZ27035	Adz27035 Adeno-sss
44	2511	86.4	734	9	ADZ27037	Adz27037 Adeno-sss
45	2506.5	86.3	735	9	ADZ27006	Adz27006 Adeno-sss

#### ALIGNMENTS

RESULT 1  
AAV71169  
ID AAV71169 standard; protein; 534 AA.  
AC AAV71169;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DB Adeno-associated virus serotype 1 capsid protein VP1.  
XX  
KM Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
KW vaccine; transgene; VP3.  
XX  
OS Adeno-associated virus 1.  
OS  
PN WC0200028061-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 02-NOV-1999; 99WC-US025694.  
XX  
PR 05-NOV-1998; 98US-0107114P.  
XX  
PA (TYPE-) UNIT PENNSYLVANIA.  
XX  
PI Wilson JM, Xiao W;  
XX WPI: 2000-376571/32.  
DR N-PSDB; AAD00772, AAD00779.  
DR  
PT Novel adeno-associated virus serotype 1 polynucleotide useful for  
XX preparation of medicament for delivery of a transgene to a host.  
XX  
PS Claim 7; Page 99-101; 109pp; English.  
XX  
CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3. The  
CC AAV-1 sequence or its fragments particularly ITRs, rep and cap coding  
CC regions, are useful in production of recombinant viral vectors for gene  
CC delivery. These vectors can be used as gene therapy vectors, vaccine  
CC vectors or antisense delivery vectors. The AAV-1 does not induce the  
CC formation of neutralising antibodies specific to any serotype of AAV  
CC hence is useful for transforming host cells, and in the preparation of a

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2005, 19:12:30 ; Search time 7.93027 Seconds  
(without alignments)  
538.765 Million cell updates/sec

Title: US-10-696-282-15

Perfect score: 3251

Sequence: 1 TAPGKRPVEQSPQPPDSSS.....NNGLYTPRRPIGTRILRPL 599

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 54001 seqs, 7132810 residues

Total number of hits satisfying chosen parameters: 54001

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:  
1: /cgn2\_6/ptodata/1/pudpa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pudpa/US06\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pudpa/US07\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pudpa/US07\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptodata/1/pudpa/US09\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pudpa/US10\_NEW\_PUB pep.\*  
7: /cgn2\_6/ptodata/1/pudpa/US11\_NEW\_PUB pep.\*  
8: /cgn2\_6/ptodata/1/pudpa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2764.5	85.0	735	US-11-184-380-24	Sequence 24, Appl
2	1830.5	56.3	598	US-10-719-311-16	Sequence 16, Appl
3	1830.5	56.3	734	US-10-719-311-4	Sequence 4, Appl
4	1700.5	52.3	588	US-11-184-380-5	Sequence 5, Appl
5	1700.5	52.3	724	US-11-184-380-4	Sequence 4, Appl
6	1690.5	52.0	544	US-10-719-311-18	Sequence 18, Appl
7	1665	53.2	532	US-11-184-380-6	Sequence 6, Appl
8	98.5	3.0	1448	US-10-485-517-212	Sequence 12, Appl
9	98.5	3.0	5935	US-10-995-561-776	Sequence 776, Appl
10	97	3.0	2504	US-10-647-956A-8	Sequence 8, Appl
11	95.5	2.9	2053	US-11-013-759-9	Sequence 9, Appl
12	94	2.9	964	US-11-089-551A-10	Sequence 30, Appl
13	93.5	2.9	1023	US-10-131-826A-200	Sequence 200, Appl
14	93.5	2.9	1116	US-10-485-517-238	Sequence 238, Appl
15	92.5	2.8	915	US-10-485-517-206	Sequence 206, Appl
16	92.5	2.8	662	US-10-647-956A-6	Sequence 6, Appl
17	91.5	2.8	662	US-11-090-439-9	Sequence 9, Appl
18	91	2.8	1466	US-11-186-284-33	Sequence 33, Appl
19	90.5	2.8	1166	US-10-821-234-964	Sequence 964, Appl
20	90	2.8	478	US-10-131-826A-138	Sequence 138, Appl
21	90	2.8	478	US-11-174-150-27	Sequence 27, Appl
22	89.5	2.8	1717	US-11-192-967-2	Sequence 2, Appl
23	89.5	2.8	1717	US-11-193-715-2	Sequence 23, Appl
24	89	2.7	1075	US-11-089-551A-23	Sequence 772, Appl
25	89	2.7	1538	US-10-995-561-772	Sequence 772, Appl

26	88.5	2.7	1377	US-10-821-234-1070	Sequence 1070, Appl
27	88.5	2.7	3623	US-10-995-561-593	Sequence 593, Appl
28	88	2.7	1263	US-10-485-517-127	Sequence 127, Appl
29	87.5	2.7	513	US-10-650-326B-16	Sequence 16, Appl
30	87.5	2.7	736	US-10-467-557-4780	Sequence 4780, Appl
31	87.5	2.7	1366	US-10-821-234-1431	Sequence 1431, Appl
32	87.5	2.7	1366	US-11-186-284-31	Sequence 31, Appl
33	87	2.7	703	US-10-467-557-1158	Sequence 7158, Appl
34	86.5	2.7	513	US-11-000-463-816	Sequence 816, Appl
35	86.5	2.7	1316	US-11-091-643-4	Sequence 4, Appl
36	86	2.6	875	US-10-933-025-12	Sequence 12, Appl
37	86	2.6	914	US-10-933-025-18	Sequence 18, Appl
38	86	2.6	914	US-11-108-172-1066	Sequence 1066, Appl
39	86	2.6	914	US-11-148-108-41	Sequence 41, Appl
40	85.5	2.6	886	US-10-873-528-126	Sequence 126, Appl
41	85	2.6	477	US-11-089-551A-34	Sequence 34, Appl
42	85	2.6	915	US-11-144-987-16	Sequence 16, Appl
43	85	2.6	1874	US-10-821-234-1182	Sequence 1182, Appl
44	84.5	2.6	488	US-10-984-376-1	Sequence 1, Appl
45	84.5	2.6	488	US-10-984-376-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-11-184-380-24  
; Sequence 24, Application US/11184380  
; Publication No. US20050255089A1  
; GENERAL INFORMATION:  
; APPLICANT: Chiorini, John  
; APPLICANT: Kotin, Robert M.  
; TITLE OF INVENTION: AAVS NUCLEIC ACIDS  
; FILE REFERENCE: 14014.032303  
; CURRENT APPLICATION NUMBER: US/11/184.380  
; CURRENT FILING DATE: 2005-07-19  
; PRIOR APPLICATION NUMBER: PCT/US99/11958  
; PRIOR FILING DATE: 1999-05-28  
; PRIOR APPLICATION NUMBER: 60/087,029  
; PRIOR FILING DATE: 1998-05-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 24  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence; Note =  
US-11-184-380-24  
Query Match 85.0%; Score 2764.5; DB 7; Length 735;  
Best Local Similarity 82.8%; Pred. No. 1.4e-214;  
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;  
QY 1 TAPGKRPVEQSPQPPDSSSGIGTGOQPAKRLINFGCTGDSSEVPPDPGLGEPAPAA 60  
DB 128 TAPGKRPVEQSPQPPDSSSGIGTGOQPAKRLINFGCTGDSSEVPPDPGLGEPAPAA 60  
QY 61 VGPTMASGGAPADNNEGADGAGNAGNCHSTLWGDVITTSRTMALPYNNHLY 120  
DB 198 LGTNMATGSPAPADNNEGADGAGNAGNCHSTLWGDVITTSRTMALPYNNHLY 120  
QY 121 KOISSASTGASNDNHYFGYSTPMGYFPDFNRHCHFSRDMQRLINNMGFRPKRLNFKLF 180  
DB 258 KOISSAS-GASNDNHYFGYSTPMGYFPDFNRHCHFSRDMQRLINNMGFRPKRLNFKLF 180  
QY 181 NIQVEKVTNDGVTIANNLSTYQVPSDSSTYQLPYVLGSAHQOCLPPFPADVMIQYX 240  
DB 317 NIQVEKVTNDGVTIANNLSTYQVPSDSSTYQLPYVLGSAHQOCLPPFPADVMIQYX 240  
QY 241 YLTANSGQAVGRSSPYCLEYFPQOMLRTGNNFTFSEYVPHSSVAHSQSLDRLMNP 300  
DB 317 YLTANSGQAVGRSSPYCLEYFPQOMLRTGNNFTFSEYVPHSSVAHSQSLDRLMNP 300





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2005, 18:55:18 ; Search time 156.491 Seconds  
(without alignments)  
2700.553 Million cell updates/sec

Title: US-10-696-282-15

Perfect score: 3251

Sequence: 1 TAPGKKRPVEQSPQEPDSSS.....NNGLYTEPRPIGTRYLRPL 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: UniProt 05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3251	100.0	736	2	Q9WB88_VIRU
2	3230	99.4	736	2	Q6JC08_VIRU
3	3230	99.4	736	2	Q6JC12_VIRU
4	3229	99.3	736	2	Q6JC17_VIRU
5	3203	98.5	736	2	Q6JC10_VIRU
6	3163.5	97.3	737	2	Q6JC13_VIRU
7	2832	87.1	736	2	Q6JC19_VIRU
8	2831.5	87.1	735	2	Q6JC08_VIRU
9	2821.5	86.8	735	2	Q6JC28_VIRU
10	2820.5	86.7	735	2	Q6JC24_VIRU
11	2818.5	86.7	735	2	Q6JC21_VIRU
12	2818.5	86.7	735	2	Q6JC27_VIRU
13	2815	86.6	735	2	Q6JC31_VIRU
14	2808.5	86.4	735	2	Q6JC17_VIRU
15	2807.5	86.4	735	2	Q6JC38_VIRU
16	2807.5	86.4	735	2	Q6JC42_VIRU
17	2803.5	86.2	735	2	Q6JC44_VIRU
18	2802.5	86.2	735	2	Q6JC36_VIRU
19	2800	86.1	734	2	Q6JC02_VIRU
20	2798	86.1	734	2	Q6JC04_VIRU
21	2796.5	86.0	735	2	Q6JC26_VIRU
22	2789.5	85.8	735	2	Q6JC43_VIRU
23	2789.5	85.7	735	2	Q6JC20_VIRU
24	2784.5	85.7	735	2	Q6JC25_VIRU
25	2782.5	85.6	735	2	Q6JC04_VIRU
26	2781	85.5	734	2	Q6JC03_VIRU
27	2778.5	85.5	735	2	Q6JC08_VIRU
28	2777.5	85.4	735	2	Q6JC06_VIRU
29	2775.5	85.4	735	2	Q6JC08_VIRU
30	2775.5	85.4	735	2	Q6JC41_VIRU
31	2773.5	85.3	735	2	Q6JC07_VIRU

32	2773.5	85.3	735	2	Q6JC29_VIRU	Q6JC29 adeno-assoC
33	2771.5	85.3	735	2	Q6JC39_VIRU	Q6JC39 adeno-assoC
34	2769.5	85.2	735	2	Q6JC08_VIRU	Q6JC08 adeno-assoC
35	2768.5	85.2	735	2	Q6JC25_VIRU	Q6JC25 adeno-assoC
36	2768.5	85.1	735	2	Q6JC25_VIRU	Q6JC25 adeno-assoC
37	2766.5	85.1	736	2	Q6JC24_VIRU	Q6JC24 adeno-assoC
38	2764.5	85.0	735	2	Q6JC31_VIRU	Q6JC31 adeno-assoC
39	2764.5	85.0	735	2	Q6JC01_VIRU	Q6JC01 adeno-assoC
40	2763.5	85.0	735	2	Q6JC35_VIRU	Q6JC35 adeno-assoC
41	2763.5	85.0	735	2	Q6JC35_VIRU	Q6JC35 adeno-assoC
42	2761.5	84.9	735	2	Q6JC26_VIRU	Q6JC26 adeno-assoC
43	2759.5	84.9	735	2	Q6JC26_VIRU	Q6JC26 adeno-assoC
44	2759.5	84.9	735	2	Q6JC26_VIRU	Q6JC26 adeno-assoC
45	2758.5	84.9	735	2	Q6JC26_VIRU	Q6JC26 adeno-assoC

## ALIGNMENTS

RESULT 1	ID	Q9WB88_VIRU	PRELIMINARY;	PRT;	736 AA.
AC	Q9WB88;	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)				
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)				
DE	Capsid protein.				
OS	Adeno-associated virus 1.				
OC	Virusae; ssDNA virusae; Parvoviridae; Parvovirinae; Dependovirus.				
OX	NCBI_TaxID=85106;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RX	MEDLINE=98214338; PubMed=10196295;				
RA	Xiao W., Chittumale N., Berta S.C., McCullough B., Gao G., Wilson J.M.;				
RT	"Gene therapy vectors based on adeno-associated virus type 1.";				
RJ	J. Virol. 73:3994-4003(1999).				
DR	EMBL; AF063497; AAD27757.1; -; Genomic_DNA.				
DR	SMR; Q9WB88; 217-736.				
DR	GO; GO:0019028; C:viral capsid; IEA.				
DR	GO; GO:0005198; F:structural molecule activity; IEA.				
DR	InterPro; IPR001403; Parvo_coat.				
DR	Pfam; PF00740; Parvo_coat; 1.				
SQ	SEQUENCE 736 AA; 81375 MW; CFA8FB9B5CD0595 CRC64;				
Query Match	100.0%; Score 3251; DB 2; Length 736;				
Best Local Similarity	100.0%; Pred. No. 5.1e-216;				
Matches	599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	TAPGKKRPVEQSPQEPDSSSGIGKTGOOPAKKRLNFGQTDSESVDPQPLGEPPTAPAA	60		
DB	138	TAPGKKRPVEQSPQEPDSSSGIGKTGOOPAKKRLNFGQTDSESVDPQPLGEPPTAPAA	197		
QY	61	VGPFTMASGGAPADNNEGADGVNAGNWHGDSITLGDVITITSTRTMALPTYNHLY	120		
DB	198	VGPFTMASGGAPADNNEGADGVNAGNWHGDSITLGDVITITSTRTMALPTYNHLY	257		
QY	121	KQISSASTGASNDNHYFGISTPKNYFPDNRHCHFSPPDWRLINNMGFFPKLNFPLF	180		
DB	258	KQISSASTGASNDNHYFGISTPKNYFPDNRHCHFSPPDWRLINNMGFFPKLNFPLF	317		
QY	181	NIQVSEVTNNGVTTIANNTSTVOVFDSFYQPYVLGSAHQGCLPPFPADVMIPOYG	240		
DB	318	NIQVSEVTNNGVTTIANNTSTVOVFDSFYQPYVLGSAHQGCLPPFPADVMIPOYG	377		
QY	241	YLLTNGSQAAGRSPFCLEYFPQMLRTGNNFTFSYTFEEVPHSSYAHQSILDRMLNP	300		
DB	378	YLLTNGSQAAGRSPFCLEYFPQMLRTGNNFTFSYTFEEVPHSSYAHQSILDRMLNP	437		
QY	301	LIDQYLTYLNTQONQSGAQNKDILFSRGSFAGNSVQPKMLBPQCTROQVSKTKTDNN	360		
DB	438	LIDQYLTYLNTQONQSGAQNKDILFSRGSFAGNSVQPKMLBPQCTROQVSKTKTDNN	497		
QY	361	NSNFWTASAKYNLNGRGSINPGTAMASHHDDDKFFPMSGWMVFGKESAGASNTALDN	420		

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2005, 19:04:49 ; Search time 28.549 Seconds  
(without alignments)  
2018.770 Million cell updates/sec

Title: US-10-696-282-15  
Perfect score: 3.51  
Sequence: 1 TAPKRRPQSPQSPDSSS.....NNGLYTEPRPIGRTYRLPL 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2177.5	67.0	504	1 VCPV3A	coat protein - ade
2	1701	52.3	732	2 S52210	coat protein VP1 -
3	556	17.1	673	1 VCPVB5	coat protein VP1 -
4	495	15.2	781	1 VCPV19	coat protein VP1 -
5	258.5	8.0	723	1 VCPVPP	coat protein VP1 -
6	254.5	7.8	729	1 A60006	coat protein VP1 -
7	254.5	7.8	729	1 VCPVNA	coat protein VP1 -
8	244	7.5	722	1 VCPVW2	coat protein VP1 -
9	234	7.2	587	1 B44276	coat protein VP1 -
10	216	6.6	716	1 VCPV2M	coat protein VP1 -
11	216	6.6	727	1 VCPV1F	coat protein VP1 -
12	214.5	6.6	718	1 VCPVIM	coat protein VP1 -
13	214	6.6	722	1 VCPVME	coat protein VP1 -
14	209	6.4	727	1 VCPVFP	coat protein VP1 -
15	208	6.4	737	1 VCPVCP	coat protein VP1 -
16	204	6.3	748	1 VCPVCP	coat protein VP1 -
17	198	6.1	584	2 S45994	capid protein VP2
18	192	5.9	722	1 VCPVCN	coat protein VP1 -
19	191	5.9	702	1 VCPVAP	coat protein VP1 -
20	143.5	4.4	648	2 S50856	whn protein - rat
21	139	4.0	1072	2 A86827	hypothetical prote
22	123.5	3.8	690	2 AB0124	probable tona-depe
23	123	3.8	1142	2 T37455	enamelin precursor
24	122.5	3.8	880	1 S9BSVS	valine-tRNA ligase
25	121	3.7	635	2 P96660	protein F2K1.10 l
26	121	3.7	1113	2 S28925	nuclear pore compl
27	120.5	3.7	931	2 T49710	related to glucan
28	119	3.7	667	2 A41311	transcription fac
29	118.5	3.6	1180	2 E86719	hypothetical prote

30	118	3.6	739	2 T52289	probable transket
31	116.5	3.6	642	1 S34416	transcription fac
32	116.5	3.6	1338	2 T30565	MAP kinase kinase
33	115.5	3.6	345	1 G97024	probable phosphoe
34	114	3.5	956	2 T08144	myosinase-binding
35	114	3.5	2271	2 F90073	hypothetical prote
36	113.5	3.5	2817	2 B97033	uncharacterized pr
37	113	3.5	1071	2 E85343	hypothetical prote
38	112.5	3.5	666	2 T52648	class A helix-loop
39	112.5	3.5	3078	2 T28432	variant-specific s
40	111	3.4	1777	2 T34369	hypothetical prote
41	110.5	3.4	655	1 ALKBG	cycloallodextrin
42	110.5	3.4	1742	2 T17120	cellulase (EC 3.2.
43	110	3.4	868	2 A82515	conserved hypotet
44	110	3.4	1296	2 C82521	hemolysin-type cal
45	110	3.4	2825	2 T14271	DocA protein, stre

## ALIGNMENTS

## RESULT 1

VCPV3A

coat protein - adeno-associated virus type 2

C/Species: adeno-associated virus type 2

C/Date: 05-Apr-1983 #sequence\_rev100 05-Apr-1983 #text\_change 09-Jul-2004

C/Accession: A03698

R/Srivastava, A.; Lueby, E.W.; Berns, K.I.

J. Virol. 45, 555-564, 1983

A/Title: Nucleotide sequence and organization of the adeno-associated virus 2 genome.

A/Reference number: A03694; MUID: 83164299; PMID: 6300419

A/Status: translation not shown

A/Molecule type: DNA

A/Residues: 1-504 <SRI>

A/Cross-references: UNIPROT:P03135; UNIPARC:UPI0000127C8E; EMBL:J01901; NID:g209616; PII

A/Keyfamily: adeno-associated virus coat protein

C/Keywords: coat protein

Query Match 67.0%; Score 2177.5; DB 1; Length 504;  
Best Local Similarity 82.5%; Pred. No. 2.3e-141;  
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;

QY	66	MASGGAPMADNNEGADGVNASGNHCDSTWLDGDRVITTSRTWALPTNNHLYKQISS	125
DB	1	MATSGAPMADNNEGADGVNSGNHCDSTWMDGDRVITTSRTWALPTNNHLYKQISS	60
QY	126	ASTGASNDNHYFGYSTPFGYDFNRFHCHESPPDMQRLINNNMGRPRRLNFKLFIQVYK	185
DB	61	QS-GASNDNHYFGYSTPFGYDFNRFHCHESPPDMQRLINNNMGRPRRLNFKLFIQVYK	119
QY	186	EYTNIDGVTTIANNLSTVQVFSDEXYOLPYVLGSAHOGCLPPPADVFMIPQGYLTIN	245
DB	120	EYTNIDGVTTIANNLSTVQVFTDSEYOLPYVLGSAHOGCLPPPADVFMIPQGYLTIN	179
QY	246	NGSQAVGRSFFCYCLEYFSPQMLRTGNNFTSYTEEVFHSVYHSQSLDLRLNPLIDQY	305
DB	180	NGSQAVGRSFFCYCLEYFSPQMLRTGNNFTSYTEEVFHSVYHSQSLDLRLNPLIDQY	239
QY	306	LYYLNRTQNGSQAONKLLFSRGSBAGMSVQPKWLPQCYROORVSKTNDNNNSFT	365
DB	240	LYYLSRTVPSGTTQSLQSLQSGASIRQSRNWLPGCYROORVSKTNDNNNSFTS	299
QY	366	WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIPGKESAGASNTALDNVMTD	425
DB	300	WTGAKTKHLNGRDSLVNP--AMASHKDEDEKFFPMGVMIPGKESAGASNTALDNVMTD	357
QY	426	EEBIAKTPVATERTGTVAVNPQSSSTDPATGDVHAMGALPGMTWODRVTLOQPIAKI	485
DB	358	EEBIEGTNPVATEGVSSTNLQSGNRQAATADAVTQCVLPGMWQODRVDVLOQPIAKI	417
QY	486	PHTDGHPSPMLMGFGKPNPPOLIKNTVPAANPPEFSATKPSATITQYSTG	540

2A The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
2B which is characterised by two inverted terminal repeats (ITR) and open  
2C reading frames for rep and capsid (cap) proteins. The rep reading frame  
2D encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
2E reading frame encodes three structural proteins, VP1, VP2 and VP3. The  
2F AAV-1 sequence or its fragments particularly ITRs, rep and cap coding  
2G regions, are useful in production of recombinant viral vectors for gene  
2H delivery. These vectors can be used as gene therapy vectors, vaccine  
2I vectors or antisense delivery vectors. The AAV-1 does not induce the  
2J formation of neutralising antibodies specific to any serotype of AAV  
2K hence is useful for transforming host cells, and in the preparation of